

# **FAUT-IL AVOIR PEUR DES ALGORITHMES ?**

**Philippe Reynet**

Philippe REYNET est professeur honoraire de Kedge Business School dont il fut également le directeur des systèmes d'information. Il est aujourd'hui en charge du développement d'un test standardisé de connaissances de base en gestion dans le cadre du projet g-skor.

« La technologie nous apprend que le monde change et que notre connaissance évolue, que nos certitudes deviennent incertitudes et que nos vérités se déforment, que ce qui est aujourd'hui un ordinateur devient demain un mobile et que nos catégories mentales se décalent. Elle nous oblige à apprendre l'apprentissage, et notre éducation doit apprendre à durer. ».

*Michel Serres, « Le contrat naturel », Flammarion, 1990*

Dans le domaine des sciences et techniques informatiques, « algorithme »<sup>1</sup> est le terme générique par excellence, un algorithme étant « *un ensemble de règles opératoires dont l'application permet de résoudre un problème énoncé au moyen d'un nombre fini d'opérations* »<sup>2</sup>. Ce vocable professionnel, qui était surtout utilisé par les acteurs de l'industrie du logiciel, est passé dans le langage courant puisque les algorithmes sont désormais au cœur de notre vie quotidienne. En effet, qu'il s'agisse des moteurs de recherche (Google,...), des réseaux sociaux (Youtube, Instagram,...), des supports de l'activité marchande en ligne ou bien des outils de traduction automatique, de diagnostic médical, de gestion de risques financiers et de vidéosurveillance (entre autres exemples), toutes ces applications que nous exploitons en permanence reposent sur des algorithmes. Dès lors, l'impact de ces objets technologiques est devenu tel qu'il n'est plus possible de considérer que leur fonctionnement ne concerne que des spécialistes de la programmation informatique. En fait, il est tout à fait compréhensible de ressentir une forme d'inquiétude face à leur sophistication et la vitesse de leur déploiement, surtout lorsqu'ils contribuent à la prise de décisions qui peuvent affecter la vie d'individus. Simultanément, il est essentiel d'identifier les multiples bénéfices qu'ils peuvent apporter grâce à leur aptitude à prendre des décisions en exploitant toute la connaissance disponible, alors qu'un décideur humain sera souvent limité par ses propres capacités cognitives.

Exercer une pensée critique à propos des algorithmes nécessite d'appréhender leur nature, leur portée pratique et les enjeux associés à leur développement ; c'est dans cette perspective que s'inscrit ce texte. Cependant, parce que le « monde des algorithmes » est particulièrement divers, nous avons choisi de centrer notre propos sur deux classes d'algorithmes qui jouent un rôle majeur dans nos sociétés modernes : les algorithmes de recommandation des réseaux sociaux et ceux qui sont associés à un processus de modélisation prédictive. Les premiers seront l'objet de notre première partie dans laquelle nous présenterons leur mode d'action et les principaux risques sociaux associés à leur mise en oeuvre en nous appuyant sur l'exemple de Youtube. Dans notre seconde partie, nous décrirons les grands principes de la modélisation prédictive -à finalité de classification- fondée sur des algorithmes d'apprentissage automatique (en anglais « *machine learning* »). Nous mettrons en évidence leur apport essentiel à la prise de décision humaine mais, parce que leur exploitation est jalonnée de nombreux cas d'impact négatif sur la vie d'individus, nous analyserons les biais qu'ils peuvent contenir et qui justifient l'encadrement de leur usage.

---

<sup>1</sup> Le mot vient du nom d'un mathématicien persan du 9ème siècle, Muḥammad ibn Mūsā al-Khwārizmī.

<sup>2</sup> Source : Dictionnaire Larousse.

## LES ALGORITHMES DES RÉSEAUX SOCIAUX : OUI À LA RECOMMANDATION, NON À LA PERSUASION.

### *Algorithmes de recommandation : l'exemple de Youtube*

Avec 14,3 milliards de visites par mois, 700 000 heures de vidéo visionnées chaque minute dans le monde et 30 milliards de dollars de revenus publicitaires en 2023, l'impact social et économique de Youtube est tel qu'il est particulièrement légitime de l'observer en exerçant une pensée critique.

Dès lors, parce que plus de 70% du contenu vu sur Youtube est issu de recommandations formulées par son algorithme se pose d'emblée la question de la manière dont sont élaborées ces recommandations. Du point de vue technique, comme tous les algorithmes des plateformes de médias sociaux, celui-ci présente, à la fois, des caractéristiques proches de celles d'un algorithme « classique »<sup>3</sup> exploitant des règles arithmétiques et logiques pré-définies<sup>4</sup> et des fonctionnalités relevant de la catégorie technologique « intelligence artificielle ». Problème : le public ne connaît que très sommairement le contenu de l'algorithme -qui évolue en permanence- et les algorithmes des réseaux sont créés à partir des données que les réseaux tirent de l'analyse des activités que nous y exerçons et avec l'objectif de maximiser la durée de notre présence. Car c'est bien cela dont il s'agit : l'algorithme doit contribuer à l'atteinte des objectifs commerciaux de la plateforme à travers la maximisation de l'expérience utilisateur et de sa satisfaction, lesquelles étant considérées comme le meilleur moyen de retenir le chaland afin de l'exposer aux messages promotionnels des partenaires du réseau. Les algorithmes sont donc conçus pour optimiser l'« engagement »<sup>5</sup> de l'utilisateur et maximiser sa présence et son activité dans la plateforme grâce à des recommandations personnalisées à partir de ses activités passées et des contenus visionnés ou partagés par des usagers au comportement similaire. Tout en constatant qu'il n'est pas souhaitable, afin d'éviter la manipulation, que Youtube divulgue l'intégralité de son algorithme, il est nécessaire d'en connaître les caractéristiques générales. Ainsi,

1. Tout commence par la collecte de vastes ensembles de données élaborées à partir des activités de chaque utilisateur : identification des vidéos visionnées,

---

<sup>3</sup> Autre exemple typique d'algorithme « classique » intervenant dans la vie quotidienne de chacun : la détermination du « Nutri-Score ». En effet, le score d'un produit alimentaire est obtenu à partir d'une règle logique dans laquelle différentes valeurs de données nutritionnelles du produit sont comparées à des valeurs de référence.

<sup>4</sup> Ces algorithmes déterministes sont par nature aisément susceptibles de description et de documentation.

<sup>5</sup> L'engagement de l'utilisateur se concrétise à travers la durée des sessions et le niveau d'activité (évaluation des vidéos visionnées, commentaires,...).

durée de visionnage, évaluations (« like »), commentaires plus, bien sûr, les abonnements à des chaînes.

2. Ces données sont ensuite analysées pour identifier les préférences de l'usager. Pour cela, le traitement combine les données individuelles et les données des vidéos elles-mêmes. Il s'agit de données descriptives saisies par le créateur de la vidéo (titre, description, mots-clés,...) qui sont complétées par des catégories thématiques obtenues à partir d'analyses automatiques de leur contenu (reconnaissance vocale, analyse d'images, traitement du langage naturel). Dès lors, une fois les préférences de l'usager identifiées, la collection de vidéos susceptibles de l'intéresser s'inscrit dans l'exploitation des catégories thématiques et de leurs proximités.

3. Au-delà, l'algorithme va également identifier des utilisateurs qui ont des modèles d'activité comparables à celui d'un usager particulier de manière à sélectionner des vidéos susceptibles de répondre aux attentes de celui qui est l'objet d'une recommandation. Les modèles individuels et les corrélations sont ainsi à la base de la recommandation de l'algorithme qui les exploite pour prédire quel contenu un usager donné est susceptible d'apprécier à partir de son comportement passé et du comportement d'usagers qui affichent des profils comparables.

4. La récence des vidéos apparaît comme un élément particulièrement significatif de la proposition algorithmique.

## ***Persuasion et biais algorithmiques***

Pour Petty et Brinol<sup>6</sup>, « La persuasion est partout... L'influence sociale exercée à travers la persuasion est le moyen le plus répandu et le plus civilisé du contrôle social dont disposent les gouvernements et les individus ». A l'issue de travaux conduits au Persuasive Technology Lab de l'université Stanford, B.J. Fogg<sup>7</sup> a mis en évidence comment des sites web et, d'une manière générale, des logiciels applicatifs peuvent être utilisés pour modifier des attitudes et des comportements individuels. Pour Fogg, les ordinateurs offrent de multiples

---

<sup>6</sup> R. E. Petty & P. Brinol, Mécanismes psychologiques de la persuasion, *Diogène*, n°217, 2017.

<sup>7</sup> B.J. Fogg, *Persuasive Technology - Using computers to change what we think and do*, Morgan Kaufmann, 2003.

possibilités de persuasion (il a ainsi créé l'acronyme « captology » pour « computers as persuasive technology »). Dans le cadre du modèle du comportement humain qu'il a élaboré (fig. 1),

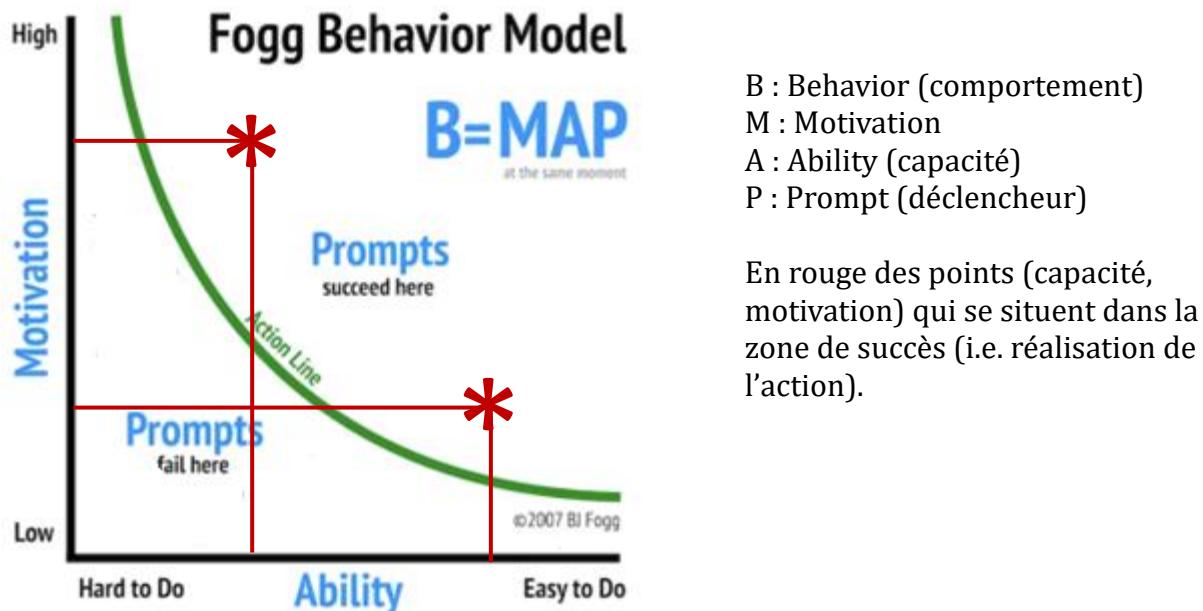


fig. 1 - Le modèle comportemental de B.J. Fogg

pour qu'un comportement se produise, il faut, simultanément,

- une motivation : plaisir/douleur, espoir/crainte, acceptation/rejet social, etc.
- une capacité déterminée par la facilité/difficulté de réalisation du comportement et qui peut être associée à l'effort nécessaire, à son coût, à la complexité cognitive, etc.
- un déclencheur, le signal qui incite à agir.

Dès lors, pour qu'un déclencheur réussisse à provoquer le comportement souhaité, il faut que les niveaux de capacité et de motivation définissent un point qui se situe à droite de la ligne verte, ce qui est notamment obtenu lorsque la motivation est très élevée (même si la réalisation de l'action est un peu difficile) ou bien lorsque, en présence d'une faible motivation, la réalisation de l'action est particulièrement aisée.

Dans le cas de l'utilisation d'un réseau social, le comportement souhaité se traduit bien souvent par un clic dont la conséquence sera le lancement ou la poursuite d'une session dans le réseau. Par exemple, une manière de motiver un utilisateur de LinkedIn à se connecter pourra consister à solliciter ses sentiments en lui envoyant le courriel suivant (le « déclencheur ») : « 10 personnes consultent votre profil - Découvrez qui a consulté votre profil ». Mais quel que soit le niveau de

motivation, la facilité de réalisation de l'action constitue un objectif permanent pour les réseaux dont l'interface doit être d'une grande simplicité d'utilisation.

Les interfaces sont au centre de l'interaction entre le fournisseur et le client. Leur design est un élément clé du succès comme le montre l'étude réalisée par des chercheurs du MIT pour BT Group (British Telecom) dans le cas d'un site web marchand<sup>8</sup> (mais dont les conclusions s'appliquent directement aux réseaux sociaux). Le but de l'étude était de tester l'influence, sur l'intention d'achat, de l'alignement (ou non) de la façon de présenter les produits sur le style cognitif des individus. Résultat : lorsque la présentation (forme, niveau de détail des caractéristiques techniques,...) est alignée sur le style cognitif (intuitif vs analytique), les intentions d'achat augmentent de 20%.

Mais la critique qui est le plus fréquemment formulée à l'égard des algorithmes des réseaux sociaux concerne les problèmes de polarisation et de contenu nuisible. Est-elle justifiée ? A l'issue d'une étude du système de recommandation de Youtube réalisée en 2023 par des chercheurs de l'Université de Californie à Davis<sup>9</sup>, ceux-ci formulent, sur la base d'analyses approfondies de données empiriques, les constats suivants :

*« Les systèmes de recommandation algorithmiques des plateformes de médias sociaux sont essentiellement conçus pour optimiser l'engagement des utilisateurs. Montrer aux gens davantage de ce qu'ils aiment est une caractéristique de tous les systèmes de recommandation et en soi, cela n'est pas alarmant. Cependant, dans le cas du contenu politique, il existe des préoccupations selon lesquelles la personnalisation et l'incitation à maximiser l'engagement des utilisateurs conduisent à une situation dans laquelle le contenu recommandé amplifie les préjugés antérieurs des utilisateurs et réduit leur exposition à des points de vue contrastants. Dans des cas extrêmes, la personnalisation et l'optimisation de l'engagement peuvent orienter certains utilisateurs vers un contenu de plus en plus radical, conspirationniste ou autrement problématique... En résumé, notre audit révèle la prévalence de recommandations bienveillantes, une augmentation limitée de l'extrémisme idéologique, et une croissance des recommandations vers des chaînes YouTube hautement problématiques. ».* A la suite de leur investigation, les auteurs de l'article suggèrent que les plateformes rendent leurs systèmes de recommandation plus transparents à leurs utilisateurs, aux scientifiques et aux instances de régulation.

<sup>8</sup> Urban, Liberali, Hauser, « Website morphing », *Journal of Marketing*, 2009

<sup>9</sup> Muhammad Haron, Magdalena Wojcieszak, Anshuman Chhabra, Xin Liu, Prasant Mohapatra et Zubair Shafiq, « Auditing YouTube's recommendation system for ideologically congenial, extreme, and problematic recommendations », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 120, n°50, 2023.

## ***Réguler et contrôler l'usage des algorithmes dans les plateformes des réseaux sociaux***

L'Union Européenne (UE) a mis en place en avril 2023 l'European Center for Algorithmic Transparency (ECAT<sup>10</sup>) pour « *fournir une expertise scientifique et technique afin de soutenir l'application de la loi sur les services numériques (Règlement sur les Services Numériques<sup>11</sup>) et pour poursuivre la recherche sur l'impact des systèmes algorithmiques déployés par les plateformes en ligne et les moteurs de recherche... L'impact sociétal croissant des plateformes en ligne telles que les réseaux sociaux, les places de marché en ligne et les moteurs de recherche a créé un besoin urgent de surveillance publique des processus au cœur de leurs activités. Les processus automatisés déployés pour modérer le contenu et organiser l'information pour les utilisateurs méritent une attention particulière, car ils affectent tout, de nos interactions sociales à notre consommation de nouvelles et de divertissement, en passant par nos habitudes d'achat.*

*À l'ECAT, notre mission consiste à contribuer à cette surveillance de deux manières principales. Tout d'abord, nous fournissons une assistance technique et des conseils pratiques pour l'application de la loi sur les services numériques. Deuxièmement, nous menons des recherches sur l'impact à long terme des systèmes algorithmiques pour éclairer l'élaboration des politiques et contribuer au débat public.*

*Tout au long de notre travail, nous adoptons une approche interdisciplinaire en intégrant des perspectives techniques, éthiques, économiques, juridiques et environnementales. Nous collaborons également avec une communauté internationale de chercheurs et de praticiens au sein de la sphère universitaire, de la société civile, des administrations publiques nationales et de l'industrie. »*

Quant au Digital Services Act qui constitue désormais le cadre juridique pour les services numériques au sein de l'EU, en réglementant la création et la diffusion de services fournis par les plateformes en ligne et en renforçant la responsabilité des entreprises qui les exploitent, il contribue à la protection des utilisateurs de ces services et à leur sécurité. En introduisant la possibilité de sanctionner le non respect de la loi par des amendes pouvant aller jusqu'à 6% du chiffre d'affaires mondial de l'entreprise concernée, il concrétise l'importance de la maîtrise du digital dans nos sociétés modernes. Cela passe par l'affirmation que « ce qui est illégal hors ligne doit aussi l'être en ligne » et se traduit, en particulier, par l'obligation de renforcement de la modération des contenus et l'interdiction du

---

<sup>10</sup> [https://algorithmic-transparency.ec.europa.eu/index\\_en](https://algorithmic-transparency.ec.europa.eu/index_en)

<sup>11</sup> <https://www.vie-publique.fr/eclairage/285115-dsa-le-reglement-sur-les-services-numeriques-ou-digital-services-act>

ciblage publicitaire sur les mineurs ou à partir de données sensibles, notamment religieuses et politiques<sup>12</sup>.

## MODÉLISATION PRÉDICTIVE : UN APPOINT ESSENTIEL, UN CADRE À ÉLABORER

Qu'il s'agisse de formuler un diagnostic médical, de déterminer si un emprunteur risque d'être un mauvais payeur ou bien de détecter un type d'émotion sur un visage (trois exemples parmi une infinité de situations comparables qui définissent autant de possibilités d'intervention d'un artefact logique), la réalisation de ces activités est de plus en plus souvent confiée à un algorithme, totalement (l'outil est alors le « décideur ») ou partiellement (« aide à la décision »).

Dans ces différents exemples, la modélisation prédictive poursuit une finalité de « classification ». En effet, il s'agit de classer un « individu » (au sens large du terme) dans une catégorie (« classe ») parmi un ensemble pré-défini de catégories (exemple : « bon payeur », « mauvais payeur »). L'algorithme effectue cette affectation -à laquelle il peut associer une probabilité de pertinence- à partir des valeurs, observées pour l'individu concerné, de caractéristiques dont on a considéré lors de la création de l'algorithme qu'elles sont les déterminants de l'appartenance à une classe (fig. 2).

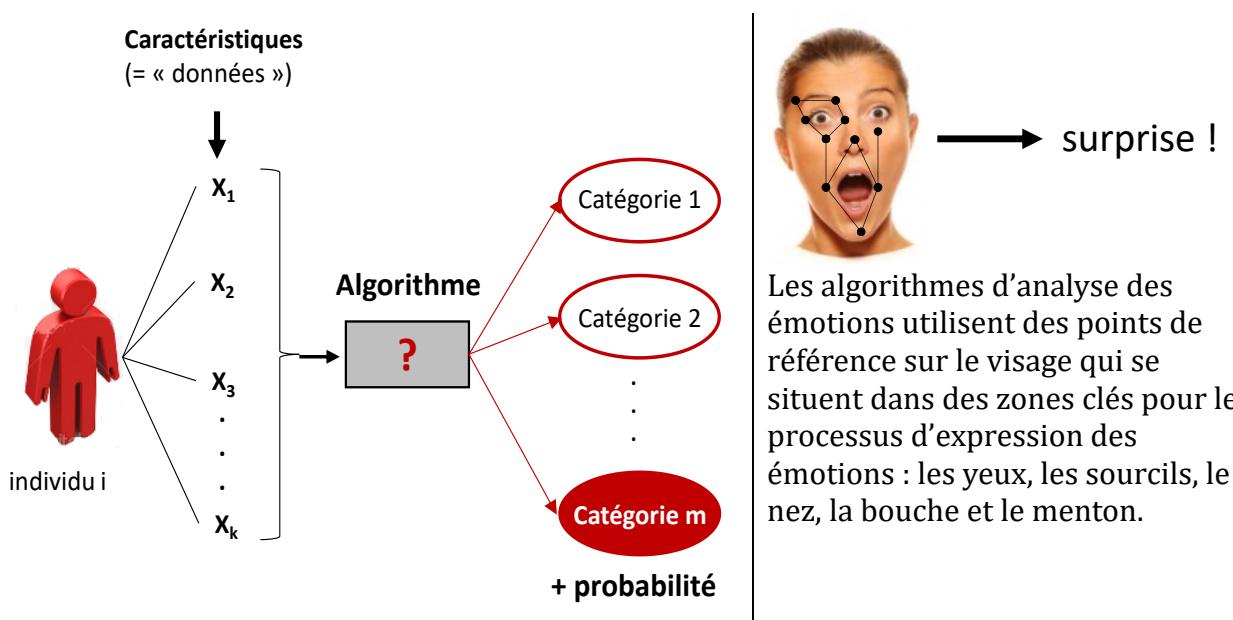


Fig. 2 - Classification d'un individu à partir d'un algorithme

<sup>12</sup> Sur ce point, on pourra consulter utilement le document suivant diffusé par l'European Data Protection Board (EDPB) : « Guidelines 8/2020 on the targeting of social media users ». url : [https://edpb.europa.eu/our-work-tools/our-documents/guidelines-82020-targeting-social-media-users\\_en](https://edpb.europa.eu/our-work-tools/our-documents/guidelines-82020-targeting-social-media-users_en)

## Comment un tel algorithme est-il élaboré ?

Les algorithmes qui sont créés pour effectuer ces prédictions n'ont qu'un lointain rapport avec les artefacts logiques de l'informatique traditionnelle. Leur aptitude à affecter un individu à une classe est le résultat de l'acquisition (« apprentissage ») d'une capacité à délimiter, dans l'espace défini par les différents facteurs d'influence, des territoires les plus homogènes possibles du point de vue de la classe dont relèvent les individus qui les composent.

Pour comprendre ce mécanisme d'apprentissage, envisageons le cas d'un vendeur de tondeuses à gazon qui se demande s'il peut prédire qu'un client a une forte probabilité d'acheter une tondeuse à gazon autoportée en fonction de la superficie de son terrain et de son revenu, caractéristiques dont il pense qu'elles sont les principaux détermi-nants de la décision d'achat. Il doit donc élaborer, à partir de données empiriques, un modèle qui relie la possession (resp. non possession) d'une tondeuse à gazon autoportée aux valeurs du couple (revenu, superficie du terrain). A l'issue d'une collecte de données auprès d'un échantillon représentatif de consommateurs, il a obtenu le graphique ci-après (fig. 3) dans lequel les possesseurs de tondeuse à gazon autoportées apparaissent en rouge et les possesseurs d'une tondeuse à gazon poussée sont représentées par un losange bleu.

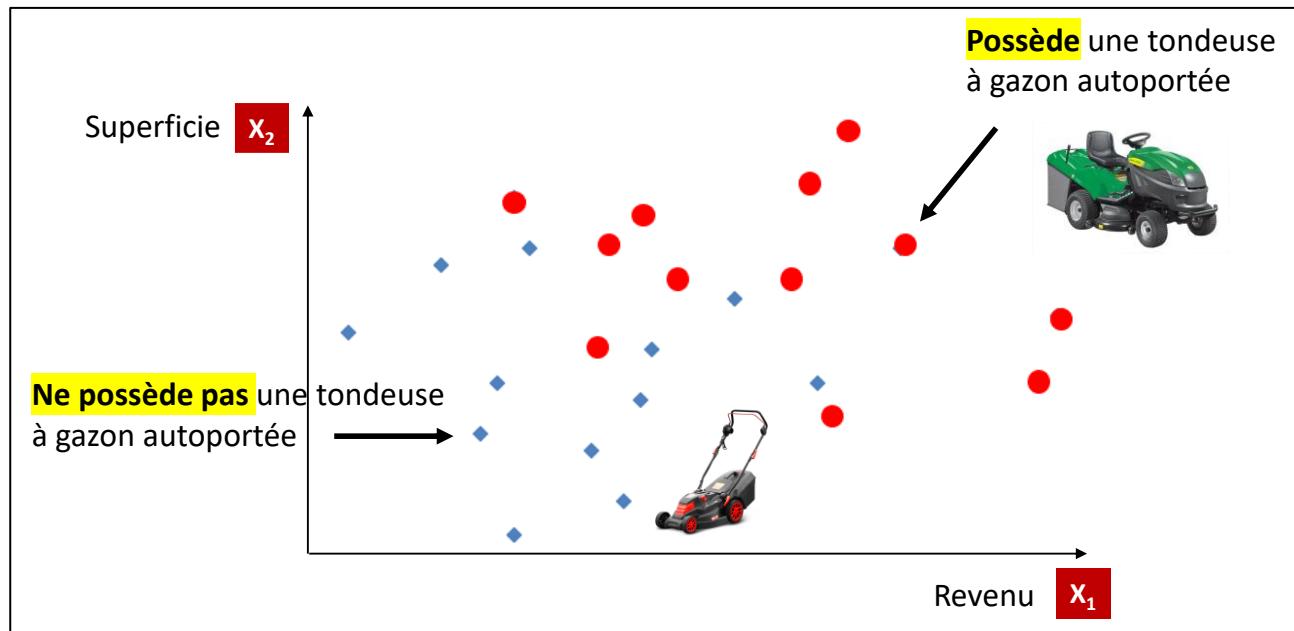
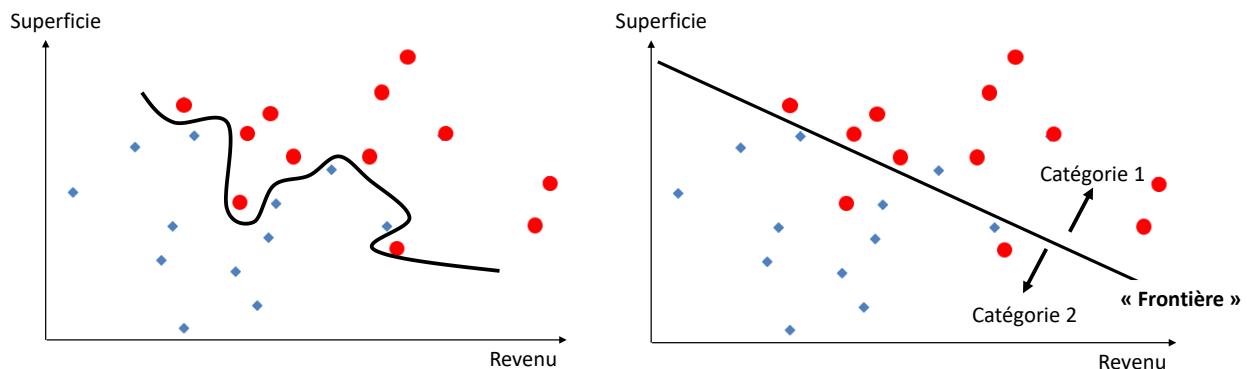


Fig. 3 - Relation entre un comportement et ses déterminants

Dans cet exemple, nous avons donc deux catégories qui définissent les résultats possibles de la prédiction et deux facteurs déterminant le comportement. Cet ensemble de données constitue l'échantillon à partir duquel on va entraîner une

structure algorithmique à base de réseau de neurones. On parle alors d' « *apprentissage supervisé* » car la classification des observations a été déjà faite manuellement afin que l'algorithme puisse d'emblée mettre en relation données initiales et classification. Dans cette phase d'entraînement, le réseau va « apprendre » (automatiquement) à séparer les individus de telle manière que le modèle décisionnel soit le plus efficace possible c'est-à-dire qu'il permette de minimiser la proportion d'individus mal classés.



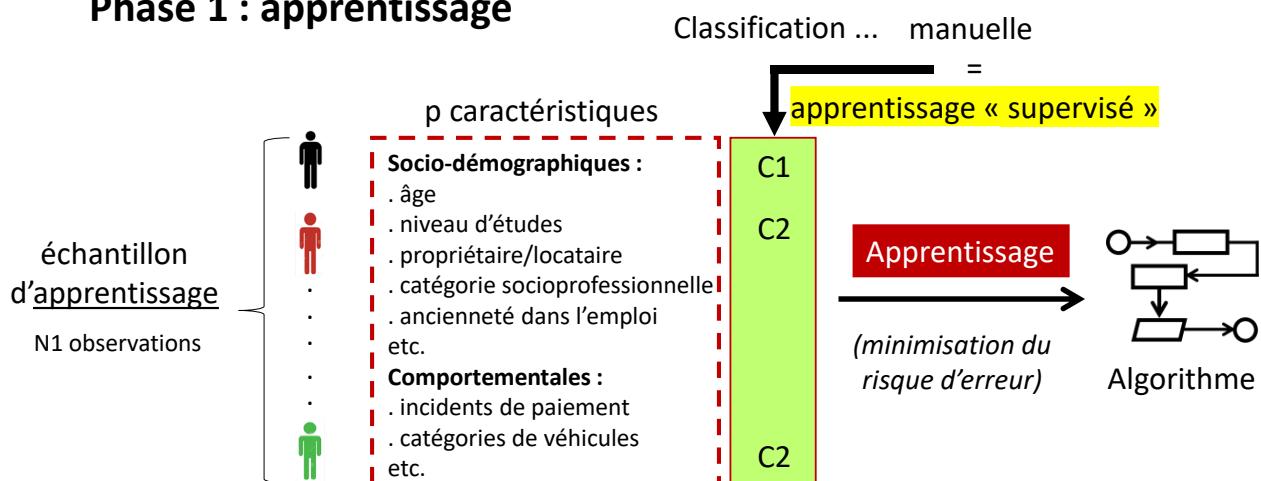
Ici, la frontière idéale entre les deux ensembles. Sa forme la rend pratiquement non modélisable mais elle n'a pas non plus un grand intérêt car elle est aussi très contextuelle et liée à l'échantillon d'apprentissage avec, potentiellement, des observations « aberrantes ».

Là, la frontière que l'on obtient avec un modèle linéaire simple. Le résultat est imparfait car il conduit à 1 mal classé bleu et 2 mal classés rouges mais il est d'un intérêt opérationnel évident. Le modèle peut sans doute être amélioré par un réseau de neurones dont les paramètres finaux sont obtenus lorsque la qualité de la séparation ne peut plus être améliorée.

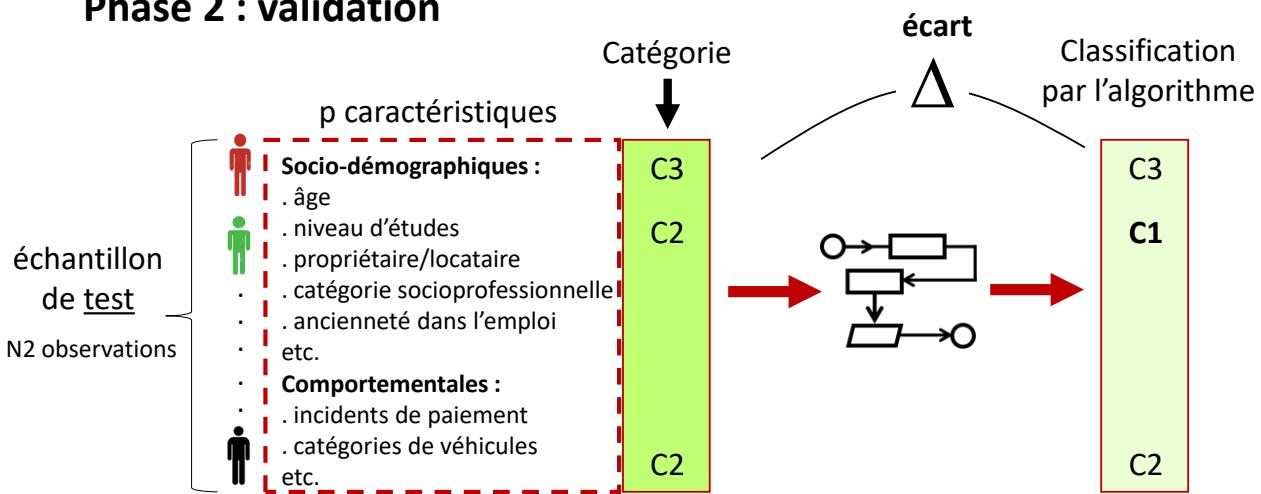
Fig. 4 - Un modèle prédictif

Dans la pratique, la classification va faire intervenir non pas deux mais  $k$  catégories et  $n$  déterminants qui vont se traduire, au sein du même ensemble de données d'entrée, par  $n$  caractéristiques hétérogènes puisque on trouvera souvent un mix de données numériques (socio-démographiques : âge ; biologiques : dosage sanguin ; etc.), catégorielles (CSP, ...), multimedia (sons, images), etc. Quant au processus de modélisation, il doit comporter deux phases : une phase d'apprentissage réalisée à partir d'un premier échantillon d'observations (c'est « l'échantillon d'apprentissage »), suivie d'une phase de validation qui va mobiliser un autre échantillon (« échantillon de test »).

## Phase 1 : apprentissage



## Phase 2 : validation



La comparaison de la classification prédite par l'algorithme aux catégories réelles des individus de ce deuxième échantillon permet d'évaluer la qualité finale de l'algorithme afin de décider, soit sa mise en exploitation, soit sa révision (sachant que, une fois en exploitation, il sera associé à un processus permanent d'amélioration).

### ***Un énorme potentiel d'aide à la décision mais des biais à contrôler***

Les algorithmes de classification issus d'un apprentissage automatique sont d'un intérêt majeur dans de multiples situations. Le diagnostic médical est un domaine privilégié car les algorithmes peuvent analyser de grandes quantités de données biologiques, génomiques, d'imagerie ou comportementales issues des dossiers médicaux des patients. Par exemple, la société française Sophia Genetics<sup>13</sup> a

<sup>13</sup> <https://www.sophiagenetics.com/>

développé un algorithme de classification de mutations génétiques qui permet de caractériser et reconnaître des variants génétiques pathogènes et porteurs de cancers ou de maladies héréditaires. Dans le domaine de la santé l'apport des algorithmes est essentiel car ils contribuent à la détection précoce des maladies, à la classification de pathologies ou à la prédition de résultats de traitements. Utilisés sous la supervision de professionnels de santé qualifiés, ils constituent dès à présent une aide à la décision considérable et bénéficient d'une amélioration continue.

Cependant, la nature même du processus d'élaboration d'un algorithme de modélisation prédictive détermine directement des possibilités de biais qui pourront engendrer des décisions à impact négatif. En effet, à l'instar du biais cognitif qui vient fausser le jugement d'un individu, le biais algorithmique est un phénomène qui altère le résultat d'un algorithme en le rendant partial, voire préjudiciable. Mais le développement des algorithmes dans un nombre croissant de domaines, par exemple pour accompagner les activités et les décisions de police, justice ou médecine, leurs biais peuvent avoir des conséquences particulièrement graves.

Parmi tous les types de biais recensés, trois nous semblent mériter une attention particulière : le « biais de données », le « biais de variable omise » et le « biais de sélection ». Le premier est liée à des insuffisances dans les données de l'échantillon d'apprentissage. Ce type de biais était par exemple présent dans une première version de l'algorithme utilisé par le recrutement d'Amazon jusqu'en 2017. En effet, l'algorithme avait tendance à favoriser les candidats masculins lors de la présélection des CV car il avait appris à valoriser certains mots-clés, expériences ou réalisations qui étaient plus fréquemment présents dans les CV d'hommes, ce qui a eu pour effet de pénaliser les candidats féminins. Un autre exemple est fourni par un algorithme de la justice américaine utilisé pour évaluer le risque de récidive. L'algorithme se trompait plus fréquemment pour les Afro-américains : 45% des accusés noirs qui n'ont pas récidivé dans les deux années étaient classés à tort dans la catégorie « risque élevé » contre 23% des blancs.

Le biais de variable omise. L'algorithme ne peut qu'exploiter les données des variables disponibles. Or, certaines variables manquantes peuvent être essentielles à la modélisation. Comme le soulignent Bertail et *al.*, « *certaines compétences humaines dites tacites sont difficiles à codifier et à incorporer dans des algorithmes de recrutement. Il s'agit de capacités telles que le leadership, l'initiative entrepreneuriale, le travail en équipe ou encore l'intelligence émotionnelle. Si ces capacités ne sont pas prises en compte, mais qu'elles sont négativement corrélées aux résultats scolaires, l'algorithme risque de pénaliser sur le marché du travail les* ».

*personnes avec des résultats scolaires moyens mais qui demeurent indispensables pour monter une équipe travaillant sur un projet innovant. »<sup>14</sup>*

Le biais de sélection. Il se réfère à la distorsion ou à l'inexactitude dans les résultats d'un algorithme, causée par la non représentativité des données d'entraînement. Ce type de biais peut influencer les décisions prises par l'algorithme, conduisant à des résultats qui favorisent indûment certains groupes ou idées au détriment d'autres. Pour illustrer cette catégorie de biais, Bertail et ses co-auteurs citent l'exemple du score de crédit : « *Pour déterminer la catégorie de risque de l'emprunteur, les algorithmes calculent un score en se basant sur les personnes qui ont été éligibles à un emprunt dans un établissement. L'algorithme utilise alors seulement les informations disponibles sur une partie de la population et ignore les dossiers de toutes les personnes à qui les banques ont refusé un prêt, celles qui n'ont jamais eu besoin d'emprunter, celles qui ont fini de rembourser leurs emprunts, et enfin celles qui ont des emprunts dans d'autres établissements.* »

Conscients des effets potentiellement néfastes des algorithmes de modélisation prédictive les pouvoirs publics des démocraties occidentales cherchent, depuis plusieurs années, à encadrer et réguler leur création et leur utilisation autour de trois enjeux majeurs : interprétabilité et explicabilité des algorithmes, transparence et auditabilité, responsabilité algorithmique. En France, EtaLab<sup>15</sup> est un département de la direction interministérielle du numérique en charge de la coordination de la conception et de la mise en œuvre de la stratégie de l'État dans le domaine de la donnée. L'application du principe de « redevabilité » aux administrations signifie que celles-ci doivent, entre autres obligations, *signaler* quand un algorithme est utilisé, le *rendre accessible* en publiant le code source et la documentation associée, *permettre la contestation* en indiquant les voies de recours possibles.

---

<sup>14</sup> Patrice Bertail, David Bounie, Stephan Cléménçon et Patrick Waelbroeck, « Algorithmes et discrimination : les biais dans les RH », HR Today, Alma Medien AG, 2020

<sup>15</sup> <https://www.etalab.gouv.fr/>

## POUR CONCLURE

Ce texte explore l'omniprésence et l'impact des algorithmes dans notre société, en mettant l'accent sur les algorithmes de recommandation des réseaux sociaux et la modélisation prédictive. Il souligne l'importance de comprendre leur nature et leur impact pour pouvoir exercer une pensée critique à leur égard. Bien que les algorithmes offrent des avantages considérables, comme l'amélioration de la prise de décision humaine, ils soulèvent également des préoccupations en matière de vie privée, de sécurité et d'éthique. Le texte appelle à une régulation et à un contrôle plus stricts pour minimiser les risques de biais et d'abus, tout en maximisant leur potentiel bénéfique. Il met en lumière la nécessité d'un équilibre entre innovation technologique et protection des individus, soulignant l'importance d'une approche éclairée et responsable de l'utilisation des algorithmes dans la société moderne.

En reconnaissant les défis et opportunités présentés par les algorithmes, nous sollicitons naturellement l'élargissement de la discussion sur l'intelligence artificielle (IA). L'IA, en tant que prolongement de la logique algorithmique, offre un potentiel encore plus grand pour transformer nos sociétés, nos économies et nos vies personnelles. Toutefois, elle soulève également des questions éthiques, sociales et techniques complexes. La nécessité d'une réglementation adaptée, d'une éthique forte et d'une participation publique informée est cruciale pour naviguer dans l'avenir de l'IA. Ce dialogue sur l'IA ne se limite pas aux experts; il doit inclure tous les acteurs de la société pour façonner un avenir où la technologie sert l'humanité tout en respectant nos valeurs fondamentales.